

# Quantum Control Landscapes: A Closer Look

Pierre de Fouquieres and Sophie G. Schirmer

Department of Applied Maths and Theoretical Physics, University of  
Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom

E-mail: plbd2@cam.ac.uk, sgs29@cam.ac.uk

**Abstract.** The control landscape for various canonical quantum control problems is considered. For the class of pure-state transfer problems, analysis of the fidelity as a functional over the unitary group reveals no suboptimal attractive critical points (traps). For the actual optimization problem over controls in  $L^2(0, T)$ , however, there are critical points for which the fidelity can assume any value in  $(0, 1)$ , critical points for which the second order analysis is inconclusive, and traps. For the class of unitary operator optimization problems analysis of the fidelity over the unitary group shows that while there are no traps over  $U(N)$ , traps already emerge when the domain is restricted to the special unitary group. The traps on the group can be eliminated by modifying the performance index, corresponding to optimization over the projective unitary group. However, again, the set of critical points for the actual optimization problem for controls in  $L^2(0, T)$  is larger and includes traps, some of which remain traps even when the target time is allowed to vary.

PACS numbers: 02.30.Yy, 03.67.-a, 03.65.Nk,

## 1. Introduction

Although quantum theory has been in existence for about a century, until recently, the main emphasis in the field was on constructing Hamiltonian models and solving the Schrodinger equation. Although it was recognized that external fields and potentials could change the Hamiltonian and thus the dynamics of a system, and such external fields were certainly used in many areas from nuclear magnetic resonance to atomic physics to effect changes to the system, it was only rather recently that the full potential of using such external fields was recognized. Since then the subject of control of quantum systems has developed from a niche area into a subject of rapidly growing interest and importance, with an ever increasing number of applications ranging from quantum chemistry to quantum information processing (see e.g. [1]). Quantum control has been applied, for instance, to influence the outcome of chemical reactions [2], to prepare entangled states [3], which are a resource in quantum metrology and information processing, and to realize quantum gates [4], which are fundamental building blocks for a quantum computer. The scope of the applications is vast.

As the potential and importance of quantum control was realized, attempts were made to formulate quantum control problems abstractly, and establish solid mathematical foundations. Some pioneering attempts were necessarily imperfect and the theory is still under development, but there have been a number of interesting theoretical results. One important area is the control landscape [5]. While it is not too difficult in many cases to formulate a particular task in terms of a control optimization problem, i.e., of finding an admissible control that maximizes a performance index, solving the resulting control optimization problems is challenging, especially for large or complicated systems, and in almost all cases they can only be solved computationally using numerical optimization strategies. Since the evaluation of the performance index for quantum control problems generally requires the solution of an operator or partial differential equation, evaluation of the latter becomes computationally demanding and efficient algorithms are therefore crucial. Such algorithms exist for finding local extrema, especially when gradients can be computed relatively cheaply compared to the computational overhead involved in evaluating the performance index, as can be achieved in our case. We are however primarily interested in global optima of that function, and this is where the “landscape” defined by the objective function becomes relevant.

If the landscape is littered with many suboptimal, locally attractive critical points then optimization strategies designed to find local extrema have a very low probability of finding a global optimum from a generic starting point. In such a situation global search strategies are usually required, involving substantially greater computational overheads. If the landscape is nice, however, i.e., if most local extrema are actually global extrema, then algorithms that find local extrema are likely to succeed in finding a global optimum and the computational overhead of global search algorithms can be avoided. The landscape for quantum control problems was initially expected to be complicated, but it was quickly realized that optimization algorithms designed to find local extrema were often very successful at finding controls for which the performance index to be optimized took values very close to its *global* optimum. This has led a number of researchers to study the issue and put forward certain arguments why this should be the case. The core argument was that, contrary to expectations, the landscape for quantum control was actually extremely simple and devoid of any traps, i.e., suboptimal locally attractive critical points.

The main objective of this paper is to revisit this topic, taking a closer look at the control landscape. Previous studies [6, 7, 8, 9, 10, 11, 12, 13, 14, 15] have almost all yielded results that only apply to regular critical points, in particular showing that these can *not* be traps for typical quantum control problems. However, it is well-known in the broader field of optimal control that these are not the only critical points [16], and that the role of such non-regular critical points requires analysis [17, 18, 19]. For quantum control problems the existence of non-regular critical points has only recently been acknowledged [20] but their properties have not been studied. In particular, it is not known if such critical points can be traps. The existence of such traps would complicate the control landscape and invalidate some general results arrived at by considering only regular points. The main objective of this paper is to show that non-regular critical points do exist for a large class of quantum control systems and that at least some of these are *suboptimal extrema* for which the performance index can take many values. The particular examples we construct to rigorously prove this claim are quite artificial, but they prove that such bad points *do* exist. For problems of practical interest it is generally extremely difficult to rigorously prove that a critical point is a trap as these problems can only be solved numerically. However, we present numerical evidence that suggests trapping behavior does occur for practical quantum optimal control problems even if such behavior appears to be relatively rare, at least for the simple systems that have been studied extensively to date. This suggests that further analysis is required to fully characterize the control landscape, including the nature of potential traps and their practical relevance, e.g., for the design of efficient algorithms.

The paper is organized as follows. In Sec. II, we formally define the problem and give the mathematical prerequisites. In Section III, we take a closer look at the landscape for pure-state transfer problems. In Sec. IV we consider the landscape for unitary operator optimization problems. Finally, in Sec. V, we present the results of a numerical investigation of the control landscape for a three-qubit gate optimization problem, before concluding with a brief discussion in Sec. VI.

## 2. Background and Mathematical Prerequisites

The focus of this paper is quantum systems subject to Hamiltonian evolution governed by the Schrodinger equation. Although the latter generally applies to states evolving over an infinite-dimensional Hilbert space, we restrict our attention to finite-dimensional systems. This is the usual context for work on control of quantum systems and motivated by the fact that in practice one is usually either dealing with a system that intrinsically has a finite number of degrees of freedom such as a collection of spins, or for which the Hilbert space can be faithfully truncated, and the controlled dynamics is restricted to finite-dimensional subspace of interest. Assuming this Hilbert space has dimension  $N$ , the evolution of the system can be described by a unitary operator  $U_f(t) \in \mathbf{U}(N)$  satisfying the Schrodinger equation

$$i\hbar\dot{U}_f(t) = H_f(t)U_f(t), \quad U_f(0) = I, \quad (1)$$

where  $I$  is the identity matrix in  $\mathbf{U}(N)$  and  $H_f(t)$  is the Hamiltonian of the system, which in our case will be control-dependent, as will be indicated by the subscript  $f$ . For simplicity the control dependence will be assumed to be linear in the (real-valued)

controls  $f_m(t)$ ,

$$H_f(t) = H_0 + \sum_{m=1}^M f_m(t) H_m, \quad (2)$$

where  $H_0$  is the system's intrinsic Hamiltonian and  $H_m$  are control interaction Hamiltonians. In the following we set  $M = 1$  unless otherwise stated, choose units such that  $\hbar = 1$ , and define the two-point propagator  $U_f(t_2, t_1) = U_f(t_2)U_f(t_1)^\dagger$ .

### 2.1. Control problems

Most typical control problems for (Hamiltonian) quantum systems fall in one of the following three categories: pure-state transfer, observable control and unitary operator implementation. The pure-state transfer problem consists of preparing the system in a desired state  $|\Psi_g\rangle$ , usually at a given time  $T$ , assuming it is initialized to some state  $|\Psi_0\rangle$  at time 0. A natural measure of success in achieving this goal is the transfer probability

$$\mathfrak{F}_P(f) = |\langle \Psi_g | \Psi(T) \rangle|^2, \quad (3)$$

where in our case  $|\Psi(T)\rangle = U_f(T)|\Psi_0\rangle$  and  $U_f(T)$  is a solution of (1). For a given Hamiltonian, initial and target states, and fixed target time  $T$ , the transfer probability depends only on the choice of control  $f$ , i.e., if the controls vary over  $L^2(0, T)$  then  $\mathfrak{F}_P : L^2(0, T) \mapsto \mathbb{R}$ . The observable control problem lies in taking a system characterized by an initial density operator  $\rho_0$  at time 0 to a state  $\rho(T)$  at time  $T$  for which the selected observable  $A$  attains its maximal possible value, i.e., we aim to maximize the expectation value

$$\mathfrak{F}_D(f) = \text{Tr}[A\rho(T)], \quad (4)$$

where  $\rho(T) = U_f(T)\rho_0 U_f(T)^\dagger$  and  $U_f(T)$  is again a solution of (1). The gate synthesis control problem consists in choreographing the system dynamics to implement the unitary gate  $V$  over the time interval  $[0, T]$ . Using the Hilbert-Schmidt (Frobenius) norm to measure distance from the target gate  $V$ , this leads to the gate fidelity

$$\mathfrak{F}_V(f) = \frac{1}{N} \Re \text{Tr}[V^\dagger U_f(T)] \quad (5)$$

as a natural performance index, where  $U_f(T)$  is a solution of (1) as before.

Broadly speaking, the objective of quantum control is to find a control  $f$  to maximize one of these performance indices, where  $f$  is allowed to vary over a large function space, while  $T$  is fixed unless stated otherwise. In practical implementations of optimization algorithms for these problems,  $f$  will always be restricted to a finite-dimensional subspace, but this subspace may be arbitrarily large, and we will be chiefly interested in the ideal case where  $f$  varies over  $L^2(0, T)$  in this paper, except in Sec. 5. So in general, we wish to find the (global) maximum of a (non-linear) functional  $\mathfrak{F} : L^2(0, T) \rightarrow \mathbb{R}$ , which in our case is either  $\mathfrak{F}_P$ ,  $\mathfrak{F}_D$ , or  $\mathfrak{F}_V$ . It will often be fruitful to think of  $\mathfrak{F}$  as a composition of the solution functional  $U_f(T) : L^2(0, T) \rightarrow \mathbf{U}(N)$  taking a control field  $f$  to the corresponding propagator at time  $T$ ,  $U(T)$ , via (1), and the fidelity on the Lie group  $\mathfrak{G} : \mathbf{U}(N) \rightarrow \mathbb{R}$ , and we shall use

$$\mathfrak{G}_P(U) = |\langle \Psi_g | U | \Psi_0 \rangle|^2, \quad \mathfrak{G}_D(U) = \text{Tr}[AU\rho_0 U^\dagger], \quad \mathfrak{G}_V(U) = \frac{1}{N} \Re \text{Tr}[V^\dagger U], \quad (6)$$

respectively, to denote the equivalents of  $\mathfrak{F}_P$ ,  $\mathfrak{F}_D$  and  $\mathfrak{F}_V$  above.

## 2.2. Controllability considerations

Whether we allow the fields  $f_m$  to range over the whole of  $L^2(0, T)$  or a smaller space such as the continuous or piecewise constant functions, we must consider whether the aforementioned control problems are feasible, i.e., whether the maximum of the performance index is actually attainable, or the question of controllability. More specifically, we are interested in whether for a given system  $H_f(t)$ , and all endpoint conditions – either  $\{|\Psi_0\rangle, |\Psi_g\rangle\}$ ,  $\{\rho_0, A\}$  or  $\{I, V\}$ , there exists a time  $T$  and an (admissible) control  $\mathbf{f}$  that solves the corresponding control problem. In general, for bilinear control systems on a Lie group, as in our case, this problem reduces to one about the Lie algebra  $\mathfrak{L}$  generated by the matrices  $\{iH_m\}_{m=0}^M$ . For pure-state control problems controllability is equivalent to  $\mathfrak{L}$  being a representation of either  $\mathfrak{sp}(N/2)$ ,  $\mathfrak{sp}(N/2) \oplus \mathfrak{u}(1)$ ,  $\mathfrak{su}(N)$  or  $\mathfrak{u}(N)$ . For mixed-state (density-operator) control problems, we require  $\mathfrak{su}(N)$  or  $\mathfrak{u}(N)$ , while the system is fully unitary operator controllable only when  $\mathfrak{L} = \mathfrak{u}(N)$  [21]. Note that the distinction between  $\mathfrak{su}(N)$  and  $\mathfrak{u}(N)$  is artificial in the quantum case as the extra dimension contributes only a global phase, which is generally not physically observable. Since it is convenient, and indeed common, to fix the final time  $T$  when carrying out control optimization, we are also interested in the stronger notion of exact-time controllability, namely, whether a system has some critical time  $T_c$  such that all endpoint conditions give solvable control problems for any  $T > T_c$ . It turns out that, neglecting global phase, these notions of controllability are equivalent to the previous three by Theorem 13 of [22].

Apart from these now standard results, we will need the following.

**Theorem 1.** *Let  $H = H_0 + f(t)H_1$  be a control system,  $\mathfrak{L} = \mathfrak{u}(N)$  or  $\mathfrak{su}(N)$  and  $G = \mathbf{U}(N)$  or  $\mathbf{SU}(N)$  be associated Lie group.*

- *If the set of all commutator expressions in  $iH_0$  and  $iH_1$  joined with  $iH_0$  and  $iH_1$  span  $\mathfrak{L}$  (Lie algebra rank condition), then there exists a maximal time  $T_{\max}$  and neighborhood  $\mathcal{N}$  of  $(H_0, H_1)$  in  $i\mathfrak{L} \times i\mathfrak{L}$  such that for all systems  $s \in \mathcal{N}$  and elements  $g \in G$ , there is a control taking  $s$  to  $g$  in time  $T < T_{\max}$ .*
- *If the set of all commutator expressions in  $iH_0$  and  $iH_1$  joined with  $iH_1$  span  $\mathfrak{L}$  (exact-time Lie algebra rank condition  $\dagger$ ), then there is a critical time  $T_c$  and neighborhood  $\mathcal{N}$  of  $(H_0, H_1)$  in  $i\mathfrak{L} \times i\mathfrak{L}$  such that for all  $s \in \mathcal{N}$ ,  $g \in G$  and  $T > T_c$ , there is a control taking  $s$  to  $g$  in time  $T$ .*

*Proof.* Assuming that our system specified by  $(H_0, H_1)$  satisfies the Lie algebra rank condition, by Theorem 1 of [22] there is a set of values  $v_1, \dots, v_n$  such that the function

$$F : (t_1, \dots, t_n) \mapsto e^{-i(H_0 + v_n H_1)t_n} \dots e^{-i(H_0 + v_1 H_1)t_1}$$

has Jacobian  $dF$  of full rank at some point  $\mathbf{t}$ . This implies that the extended function

$$F' : (A, B, t_1, \dots, t_n) \mapsto (A, B, e^{-i(A + v_n B)t_n} \dots e^{-i(A + v_1 B)t_1})$$

also has Jacobian  $dF' = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ * & * & dF \end{pmatrix}$  of full rank at the point  $(H_0, H_1, \mathbf{t})$ . Thus by the inverse function theorem, choosing neighborhoods  $\mathcal{N}$  of  $(H_0, H_1)$  and  $V$  of  $\mathbf{t}$  bounded, there is a neighborhood  $W$  of  $F(\mathbf{t})$  such that  $F'(\{A\} \times \{B\} \times V)$  includes  $\{A\} \times \{B\} \times W$  for every  $(A, B) \in \mathcal{N}$ . By compactness (and connectedness), some power  $W^k$  covers

$\dagger$  Note again that the exact-time and usual Lie algebra conditions match for  $\mathfrak{su}(N)$ .

the entire group  $G$ . Letting  $T$  be the supremum of  $\sum_n t_n$  over  $V$ , we see that for all systems in  $\mathcal{N}$  the entire group is reachable in less than time  $kT$ .

Assuming  $(H_0, H_1)$  satisfies the exact-time Lie algebra rank condition, the same argument may be repeated with

$$F : (t_0, \dots, t_n) \mapsto e^{-i(H_0+v_n H_1)t_n} \dots e^{-i(H_0+v_0 H_1)t_0}$$

as constructed in the proof of Theorem 3 in [22], where now  $F$  has domain restricted to  $\sum_n t_n = T$ . This alteration shows that the critical time  $T_c$  of exact-time controllability can indeed be chosen uniformly about  $(H_0, H_1)$ , as claimed.  $\square$

### 2.3. Gradient and Hessian formulas

Finding the global optimum of a function is generally a very difficult task. On the other hand, there are many efficient algorithms to find local extrema, i.e., attractive critical points of a function. For this reason we are particularly interested in the nature of the critical points of the various functionals  $\mathfrak{F}$ , which depends only on certain local properties of the solution functional  $U_f(t)$ . To study the latter, we need the identity

$$U_{f+\Delta f}(t) - U_f(t) = -i \int_{t_0}^t U_f(t, \tau) [H_{f+\Delta f}(\tau) - H_f(\tau)] U_{f+\Delta f}(\tau) d\tau \quad (7)$$

which can be verified by differentiating both sides, and used to derive

$$\begin{aligned} U_{f+\Delta f}(T) &= U_f(T) - i \int_0^T U_f(T, \tau) \Delta H(\tau) U_f(\tau) d\tau \\ &\quad - \int_0^T \int_0^\tau U_f(T, \tau) \Delta H(\tau) U_f(\tau, \sigma) \Delta H(\sigma) U_f(\sigma) d\sigma d\tau + O(\|\Delta f\|^3) \end{aligned} \quad (8)$$

where  $\Delta H(\tau) = H_{f+\Delta f}(\tau) - H_f(\tau)$  is just  $\Delta f(\tau) H_1$  in the  $M = 1$  case. For  $\alpha, \beta \in L^2(0, T)$  we define the linear map

$$\alpha \mapsto \int_0^T \frac{\delta U_f(T)}{\delta f(\tau)} \alpha(\tau) d\tau, \quad \frac{\delta U_f(T)}{\delta f(\tau)} = -i U_f(T, \tau) H_1 U_f(\tau), \quad (9)$$

corresponding to the gradient of  $U_f(T)$  at  $f$ , and the bilinear map

$$(\alpha, \beta) \mapsto - \int_0^T \int_0^\tau U_f(T, \tau) \Delta H(\tau) U_f(\tau, \sigma) \Delta H(\sigma) U_f(\sigma) [\alpha(\tau) \beta(\sigma) + \beta(\tau) \alpha(\sigma)] d\sigma d\tau \quad (10)$$

corresponding to its Hessian. Indeed, the second order perturbative expansion (8) is the functional analog of the Taylor expansion of a function, with the gradient and Hessian functionals corresponding to the first and second order terms, respectively. Since the Hessian is naturally a linear operator mapping  $L^2(0, T)$  to itself, it is convenient to have notation for referring directly to some such operators, along with the elements of  $L^2$  themselves. We shall use  $\Pi$  to denote the unnormalized projection onto its argument, and  $\bullet$  as a placeholder for the argument of the anonymous function it is a constituent of. For example,  $\Pi[\cos(\omega \bullet)]$  refers to the projector onto the function  $t \mapsto \cos(\omega t)$ , that is, the operator that maps each vector  $\alpha \in L^2(0, T)$  to the vector  $\int_0^T \cos(\omega \tau) \alpha(\tau) d\tau \cos(\omega \bullet)$  in  $L^2(0, T)$ .

The importance of the gradient and the Hessian functionals is that the former determines the critical points of the functional  $\mathfrak{F}$ , and the latter their nature. In particular, if the Hessian at a critical point  $f$  is negative definite, then the critical point is attractive and corresponds to a local maximum over any finite-dimensional subspace of  $L^2(0, T)$ ; similarly, if it is positive definite, the critical point is repulsive and corresponds to a local minimum over any finite-dimensional subspace, and if the Hessian at the critical point can take both positive and negative values, the critical point is a saddle.

### 3. Pure-state landscape

#### 3.1. All Critical Points over $\mathbf{SU}(N)$ or $\mathbf{U}(N)$ global extrema

As mentioned in the introduction the problem of finding a control  $f \in L^2(0, T)$  to maximize the pure-state fidelity functional (3) can be viewed a composition of finding a  $U_* \in \mathbf{U}(N)$  or  $\mathbf{SU}(N)$  that maximizes  $\mathfrak{F}_P$ , and an optimization problem of finding a control in  $L^2(0, T)$  that realizes  $U_f(T) = U_*$ . Defining  $A = |\Psi_g\rangle\langle\Psi_g|$  and  $\rho_0 = |\Psi_0\rangle\langle\Psi_0|$ , we have  $|\langle\Psi_g|U|\Psi_0\rangle|^2 = \text{Tr}[AU\rho_0U^\dagger]$ , i.e., we can rewrite the pure-state optimization problem as a general observable optimization problem. It is easy to show that a necessary and sufficient condition for  $U \in \mathbf{U}(N)$  or  $\mathbf{SU}(N)$  to be a critical point of  $\text{Tr}[AU\rho_0U^\dagger]$  is that  $A$  and  $U\rho_0U^\dagger$  commute, i.e.,  $[A, U\rho_0U^\dagger] = 0$ . In our case, as  $A$  and  $\rho_0$  (and thus  $U\rho_0U^\dagger$ ) are projectors onto pure states, there are only two types of critical points: (i)  $A$  and  $U\rho_0U^\dagger$  are projectors onto orthogonal subspaces of  $\mathcal{H}$ , in which case we have  $\text{Tr}[AU\rho_0U^\dagger] = 0$ , and (ii)  $A$  and  $U\rho_0U^\dagger$  are projectors onto the same 1D subspace of  $\mathcal{H}$ , in which case we have  $\text{Tr}[AU\rho_0U^\dagger] = 1$ . Thus, the landscape for pure-state optimization over  $\mathbf{U}(N)$  or  $\mathbf{SU}(N)$  is very simple: there are only two types of critical points, corresponding to extremal values of the fidelity, i.e., global extrema, and no saddles or other critical points for which the fidelity assumes values in  $(0, 1)$ . The critical points over  $\mathbf{SU}(N)$  or  $\mathbf{U}(N)$  are sometimes referred to as *kinematic* critical points [20].

Since then several papers have attempted to show that this result extends to the actual optimization problem over  $L^2(0, T)$ , i.e., that the fidelity  $\mathfrak{F}_P$  as a functional over  $L^2(0, T)$  only has critical points for which it achieves either its global minimum 0 or maximum 1. However, the arguments put forward in [6] and [7] rely, without rigorous justification, on the property that the solution functional  $f \mapsto U_f(T)$  is of full rank *everywhere*. [13] gives another proof but relies on the same problematic assumption. [9] argues that a certain sequence of expressions starting with equation (14) and (15) therein generate the full Lie algebra generated by  $iH_0$  and  $iH_1$ , and hence, assuming controllability, span the entire Lie algebra  $\mathfrak{su}(N)$ . But this procedure only allows one to generate specific linear combinations of commutators  $\S$ , and, although it may seem plausible, these are not guaranteed to span  $\mathfrak{su}(N)$ . For the more general density matrix control problem, [9] similarly assumes that the map  $f \mapsto U_f(T)$  is of full rank everywhere to derive that the final state  $U_f(T)\rho_0U_f(T)^\dagger$  must commute with the observable  $A$  at critical points, and hence that 0 and 1 are the only critical values for pure states. [11] makes use of the same problematic assumption, referring to [9] as evidence of its validity, to re-derive this critical point characterization, along with the result that all but the global maxima and minima are saddle points. [14] recognises

$\S$  There appears to be a mistake in equations (12-15) but our assertion still applies when the corrected expressions are used.

the possibility of singular controls  $f$ , here of  $f \mapsto \Psi_f(T)$  not being full rank for a particular control  $f$ , along with the issues this raises, yet defers analysis of such points to future work. [20] finally gives a characterization of singular controls and acknowledges that these can be critical points of the performance index that do not correspond to kinematical critical points, but considers only one example involving eigenstates of a four-level system, to conclude based on numerical simulations that singular controls do not appear to be traps for this problem.

Therefore, the question of the existence of non-kinematic critical points for the pure-state fidelity (3), i.e., critical points  $f \in L^2(0, T)$  for which  $\mathfrak{F}_P$  does not assume extremal values, remains open. We now show that such critical points do exist for all control-linear Hamiltonian systems, and while most of them cannot be expected to be attractive, one can construct examples where calculation of the Hessian at such a non-kinematic critical point shows it to be strictly negative definite. Thus, not only do non-kinematic critical points exist but they can be attractive, i.e., traps. Hessian calculations also show that the Hessian need not have finite rank at such critical points. The main implication of these examples is that landscape analysis based on considering regular points only is incomplete, and that in some cases at least, traps definitively do exist. They may be generally rare but even a single trapping point can be problematic if it has a large domain of attraction. It should be noted here that the mere existence of singular controls implies that the end-point map  $f \mapsto \Psi_f(T)$  cannot be full rank everywhere, and thus that the hypothesis on which most of the landscape results cited above are predicated, cannot be satisfied in general, so that further analysis is necessary. But this still leaves open the question whether the results are true in the absence of this hypothesis, and the significance of exhibiting examples for which traps provably exist is to show that the landscape is not universally trap-free.

### 3.2. Any critical value possible for critical points over $L^2(0, T)$

Let us start by considering, for any choice of final time  $T$ , a pure state control problem for a two-level system of the general form:

$$H_0 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad H_1 = \begin{pmatrix} c & d \\ d & c \end{pmatrix} \quad (11a)$$

$$|\Psi_g\rangle = \frac{1}{\sqrt{2}}[1, e^{i\varphi}]e^{iTH_0}, \quad |\Psi_0\rangle = \frac{1}{\sqrt{2}}[e^{i\theta}, e^{-i(\theta+\varphi)}]^T, \quad (11b)$$

where we can exclude the degenerate case  $|\Psi_g\rangle = |\Psi_0\rangle$  by choosing  $\theta \not\equiv \frac{b-a}{2}T \pmod{\pi}$ . For the field  $f \equiv 0$  which is identically zero over  $[0, T]$ , we can calculate  $\langle \Psi_g | U_0(T) | \Psi_0 \rangle = \cos(\theta)$ , so that the fidelity  $\mathfrak{F}_P$  is  $\cos^2(\theta)$ . Moreover,  $f \equiv 0$  is a critical point since the general gradient formula

$$\frac{\delta \mathfrak{F}_P}{\delta f(t)} = 2 \Im \left( \langle \Psi_g | U_f(T) U_f(t)^\dagger H_1 U_f(t) | \Psi_0 \rangle \overline{\langle \Psi_g | U_f(T) | \Psi_0 \rangle} \right)$$

in this case evaluates to

$$\Im \left( ce^{i\theta} + de^{i(a-b)t-i(\theta+\varphi)} + \text{c.c.} \right) \cos(\theta) = 0 \quad \text{for all } t$$

A value of  $\theta$  can be chosen to achieve any fidelity in the interval  $(0, 1)$ , and this can be done avoiding the degenerate case by choosing the right sign for  $\theta$  when necessary.



So we already see there is a large family of control problems admitting critical points for which the fidelity assumes any possible value but this is simply a special case of the more general fact that follows.

**Theorem 2.** *Given any bilinear control system with control Hamiltonian  $H_0 + f(t)H_1$ , any target time  $T > 0$ , and any  $F \in (0, 1)$ , there exist pairs of initial and target states  $|\Psi_0\rangle, |\Psi_g\rangle$  and a control  $f$  such that  $f$  is a critical point of the fidelity  $\mathfrak{F}_P$  with critical value  $\mathfrak{F}_P = F$ .*

*Proof.* Any bilinear control system  $H_0 + f(t)H_1$  with  $f(t) = \tilde{f}(t) + \mu$  is equivalent to one with Hamiltonians  $\tilde{H}_0 = H_0 + \mu H_1, H_1$  and control  $\tilde{f}(t)$ .

Case 1: Suppose there is a value of  $\mu \in \mathbb{R}$  such that  $\tilde{H}_0$  has degenerate eigenvalues. Choose a two-dimensional subspace  $\mathcal{S}$  of this eigenspace. Then, restricted to the subspace  $\mathcal{S}$  the Hamiltonians take the form

$$\tilde{H}_0 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \quad H_1 = \begin{pmatrix} b & 0 \\ 0 & c \end{pmatrix}. \quad (12)$$

It easy to see that  $\tilde{f} \equiv 0$  is a critical point with fidelity  $\cos^2(\theta - \varphi)$  for any  $T$  for the state transfer problem  $|\Psi_0\rangle \mapsto |\Psi_g\rangle$  on  $\mathcal{S}$  with

$$\langle \Psi_g | = [\cos(\theta), \sin(\theta)e^{-i\gamma}], \quad |\Psi_0\rangle = [\cos(\varphi), \sin(\varphi)e^{i\gamma}]^T$$

if the Hamiltonians restricted to  $\mathcal{S}$  take the form (12). Since  $\mathcal{S}$  is an invariant space of  $U(t) = e^{-it\tilde{H}_0}$ , if the initial and final states are in  $\mathcal{S}$ , the formulas for the fidelity and gradient simplify to their analogues restricted to  $\mathcal{S}$ . Therefore,  $\tilde{f} = 0$  is a critical point for the state transfer problem  $|\Psi_0\rangle \mapsto |\Psi_g\rangle$  for the original system if we embed the initial and target state into the full Hilbert space  $\mathcal{H}$  in the obvious manner.

Case 2: If the eigenvalues of  $H_0 + \mu H_1$  are always distinct then we can continuously parametrize the (unique) eigenvectors  $v^{\max}(\mu)$  and  $v_{\min}(\mu)$  corresponding to the maximal and minimal eigenvalues of  $H(\mu) = \frac{1}{|\mu|+1}(H_0 + \mu H_1)$ , respectively. Letting  $\lambda^{\max}, \lambda_{\min}$  be the maximal and minimal eigenvalues of  $H_1$ , we have the identities

$$\begin{aligned} \lim_{\mu \rightarrow -\infty} \langle v^{\max}(\mu) | (-H_1) | v^{\max}(\mu) \rangle &= -\lambda_{\min} & \lim_{\mu \rightarrow \infty} \langle v^{\max}(\mu) | H_1 | v^{\max}(\mu) \rangle &= \lambda^{\max} \\ \lim_{\mu \rightarrow -\infty} \langle v_{\min}(\mu) | (-H_1) | v_{\min}(\mu) \rangle &= -\lambda^{\max} & \lim_{\mu \rightarrow \infty} \langle v_{\min}(\mu) | H_1 | v_{\min}(\mu) \rangle &= \lambda_{\min} \end{aligned}$$

Hence by continuity we must have  $\langle v^{\max}(\mu) | H_1 | v^{\max}(\mu) \rangle = \langle v_{\min}(\mu) | H_1 | v_{\min}(\mu) \rangle$  for some (finite)  $\mu$ , unless  $\lambda^{\max} = \lambda_{\min}$ , which would mean that  $H_1$  is a multiple of the identity and so any value of  $\mu$  would do. Using this  $\mu$  leads to an effective system  $\tilde{H}_0, H_1$  which, when restricted to the subspace  $\mathcal{S}$  spanned by  $\{v^{\max}(\mu), v_{\min}(\mu)\}$ , and expressed in this basis, is precisely of the form (11). Again we can extend the initial and target states in (11) to the full Hilbert space in a trivial way so that  $\tilde{f} \equiv 0$  is a critical point with fidelity  $\cos^2 \theta$  for the resulting state transfer problem with the Hamiltonian  $\tilde{H}_0 + \tilde{f}H_1$ . Restriction to  $\mathcal{S}$  is again a well-behaved procedure since the span of each eigenvector is invariant under the time evolution generated by the corresponding Hamiltonian  $\tilde{H}_0$ .  $\square$

This theorem shows that for any control system, in particular any controllable one, there exist pairs of initial and target states for which we can achieve critical values of the fidelity in  $(0, 1)$ . However, the initial and target states which exhibit

such critical points in the proof above will not in general be eigenstates of  $H_0$ . One might wonder whether critical points for which the fidelity does not achieve extremal values might be impossible if the initial and target states are restricted to eigenstates of the system Hamiltonian  $H_0$ , as is the case in many applications, but even in this case we can find counterexamples.

**Example 1. Non-extremal critical points for eigenstate transfer problem.**

Consider the system  $H_f(t) = H_0 + f(t)H_1$  with

$$H_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 1 & \sqrt{2/3} & 0 \\ \sqrt{2/3} & 2 & \sqrt{1/3} \\ 0 & \sqrt{1/3} & 4 \end{pmatrix}$$

and the initial and target states  $|\Psi_0\rangle = [1, 0, 0]^T$ ,  $|\Psi_g\rangle = [0, 0, 1]^T$ .  $\mathfrak{F}_P$  for this system has a critical point at  $f \equiv -1$  as  $\delta\mathfrak{F}_P/\delta f = \frac{8}{9}\Im[3\cos(t) - 5] \equiv 0$  with  $\mathfrak{F}_P = \frac{8}{9}\sin^4(T/2)$ , where  $T$  is the target time. Indeed, we can even verify that this system is controllable as  $H_0$  has distinct transition frequencies and  $H_1$  is connected.

**3.3. Attractive suboptimal critical points over  $L^2(0, T)$**

We can furthermore show that the critical points above, for which the fidelity does not assume extremal values, are not necessarily saddle points either. We show this for pure state examples but the problems can be reformulated as observable optimization problems with  $A = |\Psi_g\rangle\langle\Psi_g|$  for a density matrix that is a projector onto the pure state  $|\Psi(t)\rangle$ . Hence the examples also show the existence of non-kinematic critical points for which the observable fidelity  $\mathfrak{F}_D = \text{Tr}[A\rho(T)]$  assumes non-extremal values and which are not saddle points.

**Example 2. Critical saddle points with negative semi-definite Hessian.** Let  $H_f(t) = H_0 + f(t)H_1$  with

$$H_0 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Given any target time  $T$ , choose the initial and target states

$$|\Psi_0\rangle = [\cos\phi, 0, 0, \sin\phi]^T, \quad |\Psi_g\rangle = [e^{i2T}\cos\theta, 0, 0, e^{i9T}\sin\theta]^T.$$

Then  $f \equiv 0$  is a critical point of the fidelity, since  $U_0(t) = e^{-itH_0}$  is diagonal so that  $\langle\Psi_g|U_0(T, t)H_1U_0(t)|\Psi_0\rangle = 0$ , with critical value  $\mathfrak{F}_P = \cos^2(\theta - \phi)$ . The Hessian of the fidelity at this point can be computed as

$$\begin{aligned} & -2\cos(\theta - \phi) \int_0^T \int_0^T [\cos(\theta)\cos(\phi)\cos(2(\sigma - \tau)) \\ & \quad + \sin(\theta)\sin(\phi)\cos(4(\sigma - \tau))] [\alpha(\tau)\beta(\sigma) + \beta(\tau)\alpha(\sigma)] d\sigma d\tau \end{aligned}$$

which just constitutes two instances of the  $C$  operator from Appendix A and so can more clearly be written as

$$\begin{aligned} & -2\cos(\theta - \phi) \left[ \cos\theta\cos\phi(\Pi[\cos(2\bullet)] + \Pi[\sin(2\bullet)]) \right. \\ & \quad \left. + \sin\theta\sin\phi(\Pi[\cos(4\bullet)] + \Pi[\sin(4\bullet)]) \right]. \end{aligned}$$

This expression is clearly negative semi-definite whenever  $\cos \theta \cos \phi$  and  $\sin \theta \sin \phi$  are both positive, equivalently, when  $\theta$  and  $\phi$  lie in the interior of the same quadrant. In this circumstance we are exhibiting a point  $f$  which to second order is not a saddle. For example, for  $\theta = \frac{\pi}{6}$  and  $\phi = \frac{\pi}{3}$  we obtain  $\mathfrak{F}_P = \frac{3}{4}$  and Hessian taking the form  $-\frac{3}{4} [\Pi[\cos(2\bullet)] + \Pi[\sin(2\bullet)] + \Pi[\cos(4\bullet)] + \Pi[\sin(4\bullet)]]$ . By extending (8), we can compute the third order derivative of  $\mathfrak{F}_P$  in this case to be

$$12 \cos(\theta - \varphi) \int_0^T \int_0^T \int_0^T [\cos(\theta) \sin(\varphi) \sin(4\rho + 2\sigma + \tau) - \sin(\theta) \cos(\varphi) \sin(\rho + 2\sigma + 4\tau)] \Delta f(\rho) \Delta f(\sigma) \Delta f(\tau) d\rho d\sigma d\tau$$

in direction  $\Delta f$ . When  $T \geq \pi$ , the first and second order derivatives of  $\mathfrak{F}_P$  in direction  $\gamma(t) := \begin{cases} 1 & \text{if } t < \pi \\ 0 & \text{otherwise} \end{cases}$  vanish, with the third order one being  $\frac{4}{7} \sin(2(\theta - \phi))$ . The condition already imposed on  $\theta$  and  $\phi$  implies that  $|\theta - \phi| < \frac{\pi}{2}$ , hence, unless  $\theta - \phi = 0$  thus  $\mathfrak{F}_P = 1$ , one of  $\pm \gamma$  is a direction of increase, showing that  $f$  is nonetheless a saddle point.

This example shows that we cannot rule out the existence of traps by just looking at the gradient and Hessian. In fact, we can construct examples for which we have attractive critical points with fidelity in  $(0, 1)$ , i.e., traps.

**Example 3. Non-extremal critical points with negative definite Hessian (traps).** Consider the system  $H_f(t) = H_0 + f(t)H_1$  with

$$H_0 = \begin{pmatrix} 1+\varepsilon & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & b \\ 0 & 0 & b & 0 \end{pmatrix}.$$

Taking the target time to be  $T = \frac{\pi}{\varepsilon}$ , choose the initial and target states

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}}[e^{i\theta}, 0, 0, e^{-i\theta}]^T, \quad |\Psi_g\rangle = \frac{1}{\sqrt{2}}[e^{-iT(1+\varepsilon)}, 0, 0, e^{-i2T}]^T.$$

For  $f \equiv 0$  it is easy to verify that the gradient vanishes identically as before, so this is a critical point, with fidelity  $\mathfrak{F}_P = \cos^2 \theta$ . The Hessian is

$$c\Pi[1] - \int_0^{\frac{\pi}{\varepsilon}} \int_0^{\frac{\pi}{\varepsilon}} \cos \theta [\cos \theta \cos(\varepsilon(\tau - \sigma)) - \sin \theta \sin(\varepsilon(\tau - \sigma))] [\alpha(\sigma)\beta(\tau) + \beta(\sigma)\alpha(\tau)] d\sigma d\tau$$

with  $c = -b^2 \cos^2 \theta$ , which, using Appendix A, can be rewritten as

$$a_0 \Pi[1] + \sum_{k=1}^{\infty} a_k (\Pi[\sin(2k\varepsilon\bullet)] + \Pi[\cos(2k\varepsilon\bullet)]) - \cos^2 \theta (\Pi[\sin(\varepsilon\bullet)] + \Pi[\cos(\varepsilon\bullet)])$$

with the coefficients given by  $a_0 = \frac{2}{\pi} \sin(2\theta) - b^2 \cos^2 \theta$  and  $a_k = -\frac{2 \sin(2\theta)}{\pi(4k^2 - 1)}$  for  $k > 0$ . As  $-\cos^2 \theta$  is always negative, the Hessian will be strictly negative definite if  $a_k < 0$  for all  $k$ , which is equivalent to  $b^2 \cos^2 \theta > \frac{2}{\pi} \sin(2\theta) > 0$ . With  $b = 3$ , for instance, this double inequality is satisfied for all  $\theta \in (0, \frac{\pi}{3}]$ , and this range of  $\theta$  yields fidelities

$\mathfrak{F}_P = \cos^2 \theta \in [\frac{1}{4}, 1)$ . Therefore, for all of these values of  $\theta$ , the critical point  $f \equiv 0$  corresponds to a local maximum, at least over any finite dimensional subspace, with  $\mathfrak{F}_P < 1$ . Note that this example also shows that the Hessian at a (non-kinematical) critical point need not have finite rank.

By computing the rank of the relevant Lie algebra, we can verify that  $H_0, H_1$  specifies an (exact-time) controllable system at  $\varepsilon = 0$ , so that our uniform controllability result applies. Time- $T$  controllability can therefore be ensured by choosing a sufficiently small  $\varepsilon$ , which both makes  $H_0, H_1$  lie in the neighborhood of uniform controllability and  $T = \frac{\pi}{\varepsilon}$  be larger than the minimal time required for controllability.

### 3.4. Non-constant controls

Theorem 2 and the previous examples relied on constant controls to show that critical points with non-extremal values of the fidelity exist for all systems with  $H_f(t) = H_0 + f(t)H_1$ , i.e., any choice of  $H_0$  and  $H_1$ , and that such critical points need not be saddle points but can in fact be traps. It is also interesting to consider how the critical point condition  $\delta \mathfrak{F}_P / \delta f \equiv 0$  could be satisfied for a given system for non-constant controls  $f(t)$ . Let  $\langle \Psi_B | = \overline{\langle \Psi_g | U_f(T) | \Psi_0 \rangle} \langle \Psi_g | U_f(T)$ , then the critical point condition reads

$$\Im \langle \Psi_B | U_f(t)^\dagger H_1 U_f(t) | \Psi_0 \rangle \equiv 0, \quad (13)$$

where  $U_f(t)$  must satisfy the Schrodinger equation (1), i.e., we have a Differential-Algebraic equation system. Note that since  $U_f(T)$  is of course not known in advance, we set  $\langle \Psi_B | = \overline{\langle \Psi_b | \Psi_0 \rangle} \langle \Psi_b |$  for some  $\langle \Psi_b |$  and then  $|\Psi_g\rangle = U_f(T) |\Psi_b\rangle$ , a posteriori. While the existence and uniqueness of solutions to such systems is in general not trivial to ascertain, differentiating the constraint and using the Schrodinger equation leads to a more explicit form

$$0 = \Im \langle \Psi_B | H_1 | \Psi_0 \rangle \quad (14a)$$

$$0 = \Re \langle \Psi_B | [H_0, H_1] | \Psi_0 \rangle \quad (14b)$$

$$0 \equiv \Im \langle \Psi_B | U_f(t)^\dagger [H_0 + f(t)H_1, [H_0, H_1]] U_f(t) | \Psi_0 \rangle. \quad (14c)$$

Assuming the first two equations hold, which can be thought of as a two (real) dimensional constraint on  $\langle \Psi_B |$ , the system can be solved, at least locally about  $t = 0$ , by adjoining the constraint

$$f(t) = - \frac{\Im \langle \Psi_B | U_f(t)^\dagger [H_0, [H_0, H_1]] U_f(t) | \Psi_0 \rangle}{\Im \langle \Psi_B | U_f(t)^\dagger [H_1, [H_0, H_1]] U_f(t) | \Psi_0 \rangle} \quad (15)$$

to the Schrodinger equation, under the generically true condition that the denominator does not vanish at  $t = 0$ . The additional, also generic, property that  $df/dt|_{t=0} \neq 0$  guarantees that the critical point  $f$  in question is not a constant function. As the fidelity  $\mathfrak{F}_P$  for the control  $f$  is  $|\langle \Psi_b | \Psi_0 \rangle|^2$  we can construct critical points for which the fidelity does not take extremal values. Eq. (15) can be solved numerically to find non-constant non-kinematic critical points for which the fidelity assumes any desired value, as has been done for a four-level system in [20]. While some of these critical points may be non-attractive, as appeared to be the case for the example studied in [20] there is no particular reason why *all* such critical points should be non-attractive in general. Unfortunately, it is difficult to prove this, however, due to the difficulty

involved in calculating the Hessian and proving it to be strictly negative definite given only a numerical solution of (15).

#### 4. Unitary Operator Landscape

##### 4.1. No suboptimal attractive points (traps) over $\mathbf{U}(N)$

The set of critical points of the fidelity as a map from the unitary group  $\mathbf{U}(N)$  to  $\mathbb{R}$  given by  $\mathfrak{F}_V(U) = \frac{1}{N} \Re \text{Tr}[V^\dagger U]$  is equal to all  $U$  for which  $\Re \text{Tr}[V^\dagger U A]$  vanishes for every anti-Hermitian matrix  $A \in \mathfrak{u}(N)$ . Since the bilinear map  $\mathfrak{F}_V$  is an inner product over complex matrices for which the Hermitian and anti-Hermitian matrices constitute orthogonal subspaces, the latter condition is equivalent to  $W = V^\dagger U$  being Hermitian. As  $W$  is also unitary it must be of the form  $W = P_S - P_S^\perp$ , where  $P_S$  is a projector onto a subspace  $S$  of  $\mathbb{C}^N$ , and  $P_S^\perp$  the projector onto the orthogonal complement of  $S$ . From this we see immediately that  $\text{Tr}(W)$  is equal to  $N - 2d$ , where  $d$  is the dimension of the subspace  $S$ , and thus there are  $N + 1$  critical manifolds corresponding to critical values  $1 - \frac{2d}{N}$  of  $\mathfrak{F}_V(U)$  for  $d \in \{0, \dots, N\}$ . Moreover, the second derivative of  $\mathfrak{F}_V$  is  $\frac{1}{N} \Re [V^\dagger U A^2]$ , where  $A^2$  can be any negative semi-definite matrix. Thus  $U = \pm V$ , corresponding to  $W = \pm I$  are global extrema, and all other possible  $W$  have both positive and negative eigenvalues, so that  $\Re [W A^2]$  can take any value, showing that all other critical points  $U$  are saddle points.

This characterization of the critical points on  $\mathbf{U}(N)$ , which can be traced back to at least [23], provides the motivation for the claim in [8] that the landscape for unitary operator optimization has critical points only at these values of the fidelity, and that all critical points  $U$ , except  $U = \pm V$ , which correspond to the global maximum and minimum of  $\mathfrak{F}_V(U)$ , respectively, are saddle points. As in the pure-state transfer case, to conclude this for the optimization problem over  $L^2(0, T)$  from the observations about the critical points of  $\mathfrak{F}_V$  as a functional on  $\mathbf{U}(N)$ , it is implicitly assumed that the solution functional  $U_f(T)$  is regular, in the differential geometric sense of having a Jacobian of full rank, everywhere over  $L^2(0, T)$ . [12] includes the condition that variations of the controls can be used to generate any local variation of  $U(T)$  as part of the definition of controllability. As before, the full-rank-everywhere assumption is problematic as it can not be satisfied in general. Indeed, for any constant control  $f \equiv \mu$ , the trajectory  $U_\mu(t)$  will be a linear combination of expressions  $e^{-i\lambda_j t}$ , where  $\lambda_j$  are the eigenvalues of  $H_0 + \mu H_1$ . The gradient of the solution operator at this point  $-iU_\mu(T, \tau)H_1U_\mu(\tau)$  is therefore a linear combination of functions  $e^{i(\lambda_j - \lambda_k)\tau}$  for  $j, k \in \{1, \dots, N\}$ , or equivalently of the real functions  $1, \cos((\lambda_j - \lambda_k)\tau), \sin((\lambda_j - \lambda_k)\tau)$  for  $1 \leq j < k \leq N$ , of which there are only  $N^2 - N + 1 < \dim \mathfrak{u}(N) = N^2$ . More generally the rank of the Jacobian where all controls are constants  $\mu_m$  is at most  $N^2 - N + M$  because, in the eigenbasis of  $H_0 + \sum \mu_m H_m$ , the diagonal elements of each  $U_\mu(\tau)^\dagger H_m U_\mu(\tau)$  are constant functions. The observation that constant controls are singular in the sense of not being full-rank has also been made recently in [20] but without fully considering the implications for the validity of the landscape results. Also, the application of results in [19] as suggested in [20] is problematic, as the former work, originating in sub-Riemannian geometry requires a (strictly) positive definite running costs on the controls, which serves as a regularizing term and is not present in our case, and highly undesirable as it would prevent us from ever reaching the global maximum of the actual objective function. Also properties generic over the infinite dimensional space of vector fields, such as the main results of [19] or [17], need not

hold for any instance within the finite dimensional subset of right-invariant vector fields which we restrict attention to. Thus, it is still not clear whether the fidelity  $\mathfrak{F}_V$  over  $L^2(0, T)$  has critical points that do not satisfy the criteria above, and whether these can be attractive. Before we study the fidelity over  $L^2(0, T)$ , it is useful to briefly consider the critical points of  $\mathfrak{F}_V(U)$  on the Lie group in more detail, as there are some subtleties one should be aware of; in particular when the system evolution is restricted to a subgroup of  $\mathbf{U}(N)$  such as  $\mathbf{SU}(N)$ , attractive critical points may emerge even on the group and it may be desirable to eliminate these by modifying the performance index.

#### 4.2. Attractive suboptimal critical points over $\mathbf{SU}(N)$

Many problems in quantum control involve control Hamiltonians  $H_1$  that have zero trace, i.e.  $iH_1 \in \mathfrak{su}(N)$ , in which case exact-time  $\mathbf{U}(N)$  controllability cannot hold. Indeed, we always have  $\det[U_f(t)] = e^{-iT/N \operatorname{Tr}[H_0]} = e^{i\phi}$ , regardless of the control  $f(t)$ , so the reachable set at time  $T$  is restricted to matrices in  $\mathbf{SU}(N)$  times the fixed phase  $e^{i\phi}$ . Hence, even assuming that the system is  $\mathbf{U}(N)$  controllable and  $V, T$  are chosen such that  $\det(V) = e^{i\phi}$  to make  $V$  reachable, when considering the critical points of the fidelity on the group, we should really consider the critical points of  $\mathfrak{G}_V = \frac{1}{N} \Re \operatorname{Tr}[W]$  as a functional over  $W = V^\dagger U \in \mathbf{SU}(N)$ . The critical point condition is now that  $\Re \operatorname{Tr}[WA] = 0$  for all traceless  $A \in \mathfrak{su}(N)$ , and thus that  $W$  be equal to a Hermitian matrix  $R$  plus a multiple  $i\alpha$  of the identity with  $\alpha$  real. Both this and the unitary conditions are properties of the spectrum alone, so we have in general that,  $W$  is a critical point of  $\mathfrak{G}_V$  whenever its eigenvalues are  $ie^{i\theta}$  and  $ie^{-i\theta}$  with multiplicities  $d$  and  $N-d$ , for some  $d \in \{0, \dots, \lfloor N/2 \rfloor\}$ . Finally, the unit determinant condition on  $W$  forces  $e^{i\theta}$  to equal  $\exp(i\frac{\pi N}{2(N-2d)})$  times any  $|N-2d|^{\text{th}}$  root of unity. Note that when  $N > 2$ , this set of points is larger than the set of critical points over  $\mathbf{U}(N)$  which also happen to lie in  $\mathbf{SU}(N)$ . The second derivative of  $\mathfrak{G}_V(W)$  is again  $\frac{1}{N} \Re \operatorname{Tr}[WA^2]$ , but now with  $A \in \mathfrak{su}(N)$ , and only the Hermitian part  $R$  of  $W$  contributes in this expression, since  $A^2$  is negative semi-definite. It is clear that if  $R$  has eigenvalues of both signs, implying  $N \geq 3$ , then the corresponding critical point  $W$  is a second order saddle. Otherwise  $d = 0$ , and the critical point is a local maximum or minimum if  $R$  is a positive or negative multiple of the identity, while in case  $R = 0$ , we can find curves  $\alpha ie^{Ax}$  for which  $\mathfrak{G}_V$  does not vanish to third order in  $x$ , so  $W$  is a saddle point. This characterizes the attractive critical points of  $\mathfrak{G}_V(U)$  as those  $U$  of the form  $e^{i\varphi}V$  for some  $N^{\text{th}}$  root of unity  $e^{i\varphi}$  having positive real part, and continuity of the solution operator  $U_f(T)$  then implies:

**Theorem 3.** *For any controllable system  $H = H_0 + f(t)H_1$  with  $\operatorname{Tr}(H_1) = 0$ , and any  $F = \cos(2\pi k/N)$  with  $k = 1, \dots, \lceil N/4 \rceil - 1$ , there exists an open subset  $\mathcal{N}$  of  $L^2(0, T)$  such that any monotonically increasing local optimization algorithm started with  $f \in \mathcal{N}$  will never exceed a fidelity of  $F$ .*

#### 4.3. Elimination of traps by optimization over $\mathbf{PU}(N)$

These results show that for optimization of  $\mathbf{SU}(N)$ , there are traps, i.e., attractive critical points of  $\mathfrak{G}_V(U)$  with  $\mathfrak{G}_V(U) < 1$ , and this is relevant for optimization problems over  $L^2(0, T)$  with  $\operatorname{Tr}[H_1] = 0$ . As the attractive critical points differ only by a global phase factor  $e^{i\theta}$  from the target gate  $V$ , however, and we usually do not care about the global phase of a gate, these traps are not really a problem in

practice, provided the convergence condition of the algorithm is more sophisticated than checking if  $1 - \mathfrak{F}_V(U_f(T))$  is less than a certain tolerance. However, if we do not care about the global phase of the gate  $V$ , a better choice of the performance index to be optimized would be

$$\mathfrak{G}'_V : U \mapsto \frac{1}{N^2} |\text{Tr}(V^\dagger U)|^2, \quad (16)$$

which is equivalent to optimization over the projective unitary group  $\mathbf{PU}(N)$ . To see why it is a better choice, note that the critical point condition for  $\mathfrak{G}'_V$  is  $\Re[\bar{\gamma} \text{Tr}(WA)] = 0$  for every  $A \in \mathfrak{su}(N)$  with  $W = V^\dagger U$  as before, and  $\gamma = \text{Tr}[W]$ . This is satisfied only when  $\bar{\gamma}W = \alpha iI + R$  for some real  $\alpha$  and Hermitian  $R$  with all eigenvalues of the same magnitude. The second derivative at these points is  $2\Re \text{Tr}[RA^2] + 2|\text{Tr}[WA]|^2$ , the first term can only fail to ever be positive for any  $A \in \mathfrak{su}(N)$  if  $R$  is a non-negative multiple of the identity, or  $N = 2$ . In the first case, either  $\gamma = 0$  or  $W = e^{i\theta}I$ , which corresponds to  $\mathfrak{G}'_V$  attaining its minimal or maximal values respectively. Otherwise,  $R$  has eigenvalues of opposite sign, so the second derivative's first term must vanish for every  $A \in \mathfrak{su}(N)$ , and if its second term always vanishes then  $U$  must have equal eigenvalues, thus maximizes  $\mathfrak{G}'_V$ .

#### 4.4. Non-global maxima for optimization over $L^2(0, T)$

The previous section shows that suboptimal attractive critical points do arise for optimization of  $\mathbf{SU}(N)$ , and that this case is relevant for quantum control problems, but these can easily be eliminated by changing the performance index. We now turn our attention to the existence of sub-optimal attractive critical points for the actual optimization problem of interest, i.e., the problem of maximizing the fidelity  $\mathfrak{F}_V$  over  $L^2(0, T)$ . In particular we are interested in whether there are suboptimal attractive critical points, i.e., points  $U_f(T)$  for which the gradient of  $\mathfrak{F}_V$  vanishes, the Hessian is negative definite, and  $\mathfrak{F}_V(U_f(T)) < 1$  for systems that are time- $T$  controllable.

**Example 4. Traps for unitary operation optimization problem.** Consider the control problem specified by:

$$H_0 = \begin{pmatrix} 1+\varepsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad H_1 = \begin{pmatrix} a & 1 & 0 \\ 1 & b & 1 \\ 0 & 1 & c \end{pmatrix}, \quad V^\dagger = \begin{pmatrix} e^{i\varphi} & 0 & 0 \\ 0 & ie^{i\gamma} & 0 \\ 0 & 0 & ie^{\pm i\gamma} \end{pmatrix} e^{iT H_0}$$

with  $T = \frac{\pi}{\varepsilon}$ , whose gradient at  $f \equiv 0$  can be calculated as

$$\frac{1}{3} \Im \text{Tr}(V^\dagger U_f(T) H_1) = \frac{1}{3} a \sin(\varphi) + \frac{1}{3} (b+c) \cos(\gamma)$$

and whose Hessian at the same  $f \equiv 0$  expands to

$$\begin{aligned} & -\frac{1}{3} w \Pi[1] + \frac{1}{3} x \sum_{k=1}^{\infty} \frac{4}{\pi(4k^2 - 1)} (\Pi[\sin(2k\varepsilon\bullet)] + \Pi[\cos(2k\varepsilon\bullet)]) \\ & + \frac{1}{3} y (\Pi[\cos(\bullet)] + \Pi[\sin(\bullet)]) - \frac{1}{3} z (\Pi[\cos(\varepsilon\bullet)] + \Pi[\sin(\varepsilon\bullet)]) \end{aligned}$$

where the constants  $w, x, y, z$  are defined in (17) below. This formula for the Hessian again comes from applying the results of Appendix A to the original operator expression

$$\begin{aligned} & \frac{1}{3} \int \int_{0 < \sigma < \tau < \frac{\pi}{\varepsilon}} [x \sin(\varepsilon(\sigma - \tau)) - (a^2 \cos \varphi - (b^2 \pm c^2) \sin \gamma) \\ & \quad + y \cos(\sigma - \tau) - z \cos(\varepsilon(\sigma - \tau))] [\alpha(\sigma)\beta(\tau) + \alpha(\tau)\beta(\sigma)] d\sigma d\tau \end{aligned}$$

Therefore, a sufficient set of conditions for this operator to be negative definite, thus the fidelity  $\mathfrak{F}_U$  to have a local maximum over all finite dimensional subspaces at  $f \equiv 0$ , are the constraints:

$$0 = a \sin \varphi + (b + c) \cos \gamma, \quad (17a)$$

$$0 < w = \frac{4}{\pi} (\cos \gamma - \sin \varphi) + a^2 \cos \varphi - (b^2 \pm c^2) \sin \gamma, \quad (17b)$$

$$0 > x = \cos \gamma - \sin \varphi, \quad (17c)$$

$$0 \geq y = \sin \gamma \pm \sin \varphi, \quad (17d)$$

$$0 \leq z = \cos \varphi - \sin \gamma. \quad (17e)$$

We are interested in whether these conditions, which are independent of  $\varepsilon$ , can be satisfied for some choice of the other parameters, while having  $H_0, H_1$  satisfy the (exact-time) controllability Lie algebra rank condition for  $\varepsilon = 0$ , so that by our uniform controllability result a sufficiently small choice of  $\varepsilon$  would make this system exact-time controllable at  $T$ . Indeed, several such choices of parameters exist, for example: in the  $-i$  case,  $a = 5\sqrt{2/3}$ ,  $b = 4$ ,  $c = 1$ ,  $\varphi = \frac{2\pi}{3}$ ,  $\gamma = -\frac{3\pi}{4}$  with fidelity  $\frac{1}{6} \approx 42\%$ , or in the  $+i$  case,  $a = \sqrt{2/3}$ ,  $b = -1$ ,  $c = 0$ ,  $\varphi = \frac{\pi}{3}$ ,  $\gamma = -\frac{\pi}{4}$  with a fidelity of  $\frac{1}{6} + \frac{\sqrt{2}}{3} \approx 82\%$ .

#### 4.5. Variable time non-global maxima

In typical quantum control problems, the target time  $T$  is fixed, but  $T$  can be allowed to vary. If the target time  $T$  of control problems is not fixed in advance, and is consequently allowed to vary in the corresponding optimization formulation, there is some ambiguity as to how proximity between control fields is to be measured when these lie in different spaces  $L^2(0, T)$ . We choose to resolve this difficulty by optimizing over Hamiltonians of the form  $H(t) = \ell H_0 + f(t)H_1$  with  $T$  fixed, since propagating  $H(t)$  up to time  $T$  is equivalent to propagating  $H_0 + \frac{1}{\ell} f(t/\ell)H_1$  up to time  $\ell T$ . The expressions for the gradient (9) and Hessian (10) derived earlier from the perturbative expansion can be adapted to this variable time optimization framework simply by replacing all instances of  $\Delta f H_1 = H_{f+\Delta f} - H_f$  by  $\Delta \ell H_0 + \Delta f H_1$ . In particular, at a point where  $f \equiv 0$ , we find the solution functional has gradient simplifying to

$$(\ell_a \oplus \alpha) \mapsto -i \int_0^T U_f(T, \tau) H_1 \Delta f(\tau) U_f(\tau) d\tau - iU(T)TH_0\Delta\ell \quad (18)$$

while its Hessian is the bilinear map on  $\mathbb{R} \oplus L^2(0, T)$  given by

$$\begin{aligned} (\ell_a \oplus \alpha, \ell_b \oplus \beta) \mapsto & - \int \int_{0 < \sigma < \tau < T} U_f(T, \tau) H_1 U_f(\tau, \sigma) H_1 U_f(\sigma) [\alpha(\tau)\beta(\sigma) + \beta(\tau)\alpha(\sigma)] d\sigma d\tau \\ & - U(T) \int_0^T ([U_f(\tau)^\dagger H_1 U_f(\tau), H_0] \tau + H_0 U(\tau)^\dagger H_1 U_f(\tau) T) [\alpha(\tau)\ell_b + \beta(\tau)\ell_a] d\tau \\ & - U_f(T)T^2 H_0^2 \ell_a \ell_b. \quad (19) \end{aligned}$$



**Example 5.** In this framework we consider the unitary control problem specified by:

$$H_0 = \begin{pmatrix} 1+\varepsilon & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & b & 1 & 0 \\ 0 & 1 & c & g \\ 0 & 0 & g & d \end{pmatrix},$$

$$V^\dagger = \begin{pmatrix} r & 0 & 0 & se^{i(\theta-\gamma)} \\ 0 & -i & 0 & 0 \\ 0 & 0 & e^{i\varphi} & 0 \\ -se^{i\gamma} & 0 & 0 & re^{i\theta} \end{pmatrix} e^{iTH_0}$$

with  $T = \frac{\pi}{\varepsilon}$ , where a fortiori  $s = \sqrt{1-r^2}$ , and we shall set  $b = \frac{5}{2}(1-\sqrt{3})$ ,  $c = \sqrt{3}-3$ ,  $d = 3$ ,  $g = \sqrt{\frac{3}{71}(19+12\sqrt{3})}$ ,  $r = \frac{2}{9}(3-\sqrt{3}) \approx 0.28$ ,  $\varphi = \frac{\pi}{3}$ ,  $\theta = -\frac{\pi}{3}$  with  $\gamma$  arbitrary. With these parameters, the gradient at  $f \equiv 0$ ,  $\ell = 1$  always vanishes, and the Hessian evaluates to

$$\begin{aligned} & -CT^2 d\ell^2 - 0.80 \Pi[1] - 1.60 (\Pi[\cos] + \Pi[\sin]) - 0.28 (\Pi[\cos(\varepsilon\bullet)] + \Pi[\sin(\varepsilon\bullet)]) \\ & - \sum_{k=1}^{\infty} \frac{4}{\pi(4k^2-1)} (\Pi[\sin(2k\varepsilon\bullet)] + \Pi[\cos(2k\varepsilon\bullet)]), \end{aligned}$$

which is always positive definite, since  $C$ , although dependent on  $\varepsilon$ , must be at least 3.5, and the fidelity  $\mathfrak{F}_U$  here is  $\frac{1}{48}(33-2\sqrt{3}) \approx 62\%$ .

## 5. Practical Evidence for Trapping Behavior

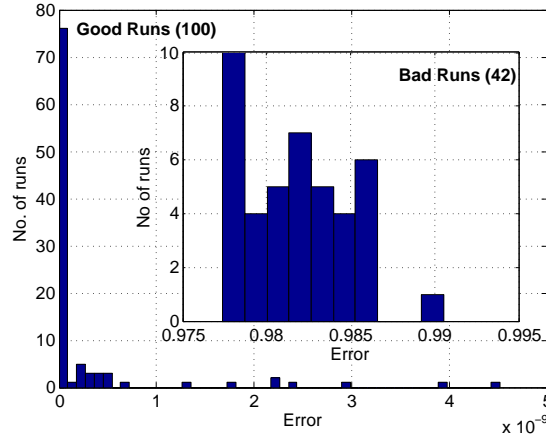
Having demonstrated the existence of suboptimal attractive critical points for quantum control optimization problems, the question arises how significant such points are in practical applications. The examples in the previous sections for which we can rigorously prove the existence of such bad points are clearly quite artificial, and mainly serve to prove a theoretical point. Unfortunately, it does not appear that any theoretical machinery is available to provide a general answer to this question for far more complex problems that are of immediately practical interest, so we will use numerical simulations to give a glimpse at the range of possibilities. For this purpose, we consider the problem of implementing a quantum Fourier transform gate for a three-qubit Ising-coupled spin chain with full local control. Explicitly, the system has the time varying Hamiltonian

$$2H(t) = Z_1 Z_2 + Z_2 Z_3 + f_1(t)X_1 + f_2(t)Y_1 + f_3(t)X_2 + f_4(t)Y_2 + f_5(t)X_3 + f_6(t)Y_3. \quad (20)$$

with six control fields  $f_1, \dots, f_6$ , where  $\{X, Y, Z, I\}$  are the unnormalized Pauli matrices and  $X_1$  is a shorthand for the tensor product  $X \otimes I \otimes I$ , etc. The quantum Fourier transform objective  $V$  is specified by

$$\langle j|V|k\rangle = \frac{\sqrt[8]{i}}{\sqrt{8}} \omega^{jk} \quad (21)$$

where  $\omega$  is the canonical choice  $e^{-i2\pi/8}$ , and the prefactor  $\sqrt[8]{i} = e^{2\pi i(m+\frac{1}{4})/8}$  for some integer  $m$  is needed to ensure  $V \in \mathbf{SU}(2^3)$  as our Hamiltonians are in  $\mathfrak{su}(2^3)$ . We

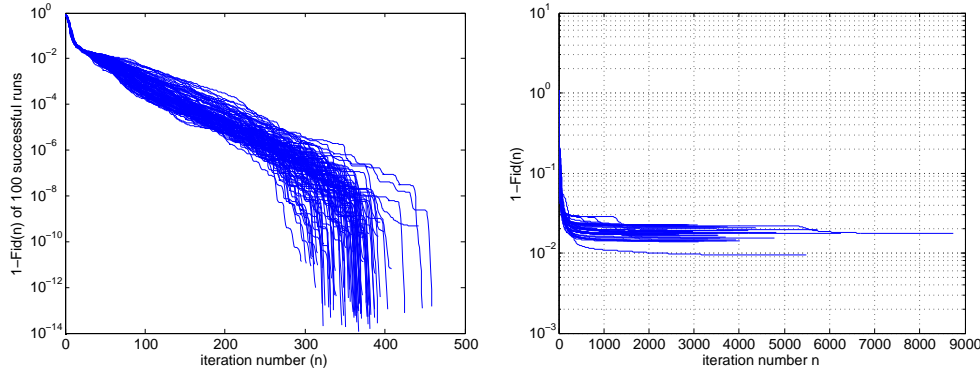


**Figure 1.** Histogram of errors at the termination for 100 selected successful runs (co-inciding with the 100 good runs plotted in Fig. 2) show that the good runs terminate with errors  $< 5 \times 10^{-9}$ , with many reaching smaller errors. The histogram of the errors for the 42 bad runs (inset) show that all but one of these terminate with errors  $\geq 10^{-2}$ , orders of magnitude larger than the terminal errors for the good runs.

choose  $m = 5$ . To solve the optimization problem numerically we use the standard parameterization of the controls  $f_r$  in terms of piecewise constant functions over  $S$  intervals forming a uniform partition of  $[0, T]$ , i.e., effectively we define  $f_r(t) = a_{rs}$  whenever  $(s-1)\frac{T}{S} < t < s\frac{T}{S}$ , and optimize over all  $a_{rs}$ ,  $s = 1, \dots, S$  variables.

This system satisfies the condition of exact-time controllability over  $\mathbf{SU}(8)$ , the case we are usually interested in, although, as is generally the case, we do not know the critical time  $T_c$  required to render the whole group accessible. Neither do we know how to ensure having a suitable number of time slices  $S$ , except that we must satisfy the dimensionality constraint  $6S \geq 8^2$ , otherwise the optimization will either get trapped or converge to unbounded fields, for almost all objectives  $V$ . However, the fact that we were always able to achieve errors  $1 - \mathfrak{F}_V$  below  $10^{-10}$  for 50 randomly selected target gates suggests that our choice of  $T = 8$  is sufficiently large, and of  $S = 140$  appropriate. Ideally, we would like control optimization to be effective even under the weaker condition that the target gate  $V$  be accessible at the given time  $T$ , using  $S$  time slices. This is certainly true, at the  $10^{-10}$  accuracy level, for our  $V$ .

For the optimization we primarily used a quasi-Newton method [24], relying on the BFGS formula to construct an approximate Hessian, and an exact analytic formula for the gradients. This choice was motivated by related work in which this procedure performed well relative to various others, in particular at high fidelities. Each run of the algorithm was initialized by setting the  $a_{rs}$  variables independently at random from a standard normal distribution, a choice which was deemed suitable considering the typical norm of high fidelity solutions. We carried out 1,000 optimization runs and found 958 of them successfully passed our pre-selected error threshold of  $10^{-4}$ , while of the remaining 42 all but one failed to reach even the  $10^{-2}$  level. A histogram of the terminal errors for 100 selected good runs versus the 42 bad runs (Fig. 1) shows that the final fidelities achieved by the good runs are many orders of magnitude higher than the terminal fidelities for the bad runs. Fig. 2 further shows that the convergence

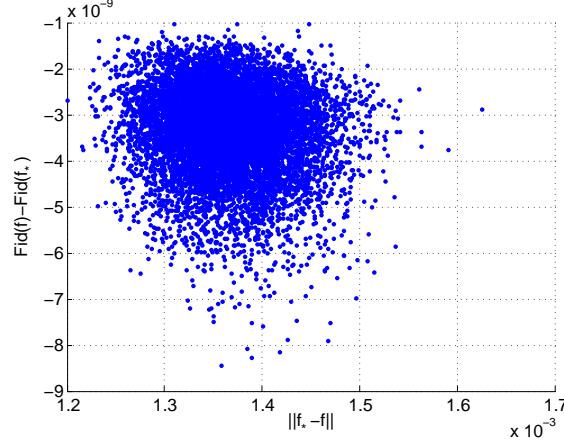


**Figure 2.** Convergence behavior for successful runs and trapped runs. Error  $1 - \mathfrak{F}_V(n)$  as a function of the iteration number  $n$  for 100 successful runs (left) shows rapid convergence to final fidelities in the range of  $10^{-10}$  and  $10^{-14}$  in less than 500 iterations, while the error for the bad runs (right) remains virtually constant over thousands of iterations at values around  $10^{-2}$ , more than 8 orders of magnitude larger than the final error for the successful runs.

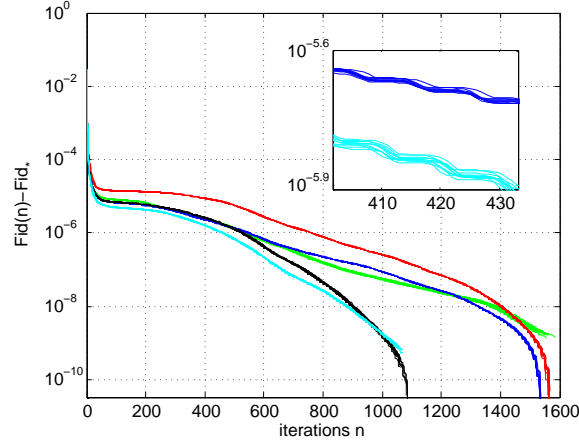
behavior for the successful runs is distinctly different from that of the failed ones. Considering that the successful runs reached fidelities  $> 99\%$  on average in a few dozen iterations taking mere seconds, the bad runs constitute effective traps in that the otherwise highly efficient quasi-Newton optimization routine gets stuck at these points and the runs are extremely time-consuming.

We further explored the behavior of the fidelity in neighborhoods of these bad points by evaluating it for 10,000 small random field perturbations around each of the presumed traps  $\mathbf{f}_*$ . The results show that the fidelities for *all* of these fields are smaller than the fidelity  $\mathfrak{F}_V(\mathbf{f}_*)$  at the presumed trap  $\mathbf{f}_*$ , suggesting that the  $\mathbf{f}_*$  are indeed in the trapping neighbourhood of a local maximum for all 42 runs that terminated at fidelities  $\mathfrak{F}_V \lesssim 0.99$ . An example of a scatter-plot of the fidelities of the perturbed fields  $\mathbf{f}$  as a function  $\|\mathbf{f} - \mathbf{f}_*\|$  in a neighborhood of trap 1 is shown in Fig. 3. We know that this is not a result of the target time  $T$  being too short or the dimension of the subspace of  $L^2$  being too small as there are solutions for the fields for  $T = 8$  and  $S = 140$  with errors  $1 - \mathfrak{F}_V(\mathbf{f})$  on the order of  $10^{-14}$ , close to machine precision. To ascertain whether these bad fields  $\mathbf{f}_*$  indeed attracted trajectories, we also restarted the optimization with perturbed fields  $\mathbf{f}^{(0)} = \mathbf{f}_* + \delta\mathbf{f}$ . Even when the size of the perturbations was increased to  $\delta f_m$  on the order of 0.1 the runs tended to remain stuck in the trapping region as illustrated in Fig. 4. Finally, we also tested other implementations, including several variants of sequential update algorithms, similar to the Krotov method *without* penalty, but starting with the bad points  $\mathbf{f}_*$  none of the runs managed to escape the trapping regions. It is difficult to ascertain if there is a single trap or possibly multiple traps in each region, and there are clearly limitations of this computational investigation, but taken together, the simulation results strongly suggest that there are suboptimal extrema and effective trapping regions, at least for the  $6 \times 140$ -dimensional search space we chose to optimize over.

It is difficult to ascertain whether these bad points would constitute traps if each of our six control fields  $f_m$  were not restricted to the 140-dimensional subspace of  $L^2(0, 8)$  consisting of piecewise constant functions. To get a glimpse of how the convergence behavior changes when we enlarge the subspace of  $L^2$  we project onto, we subdivided

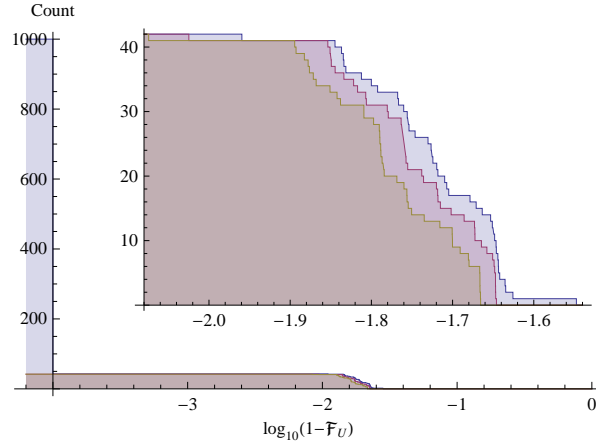


**Figure 3.** Fidelity sampled in a neighborhood of a typical trap.  $\mathfrak{F}_V(\mathbf{f}) - \mathfrak{F}_V(\mathbf{f}_*)$  is negative, bounded away from 0, for all sample points, showing that the fidelity at all of these random samples is less than the fidelity at the presumed local extremum  $\mathbf{f}_*$ , suggesting that  $\mathbf{f}_*$  is indeed a local extremum.



**Figure 4.** Convergence near five randomly selected “traps”  $\mathbf{f}_*$ . For each  $\mathbf{f}_*$ , for clarity, the *difference* of the fidelity as a function of the iteration number,  $\mathfrak{F}_V(n)$ , and the fidelity of the nearest “trap”  $\mathfrak{F}_* = \mathfrak{F}_V(\mathbf{f}_*)$  is shown for 10 runs starting with  $\mathbf{f}_* + \delta\mathbf{f}$ . For each run the difference in the fidelities vanishes, i.e., the fidelities seem to converge to the respective fidelity value for the closest trap. Inset shows close-up of a small portion to better resolve the trajectories.

the time grid three-fold to  $S = 420$ , and restarted the optimization, starting with the field  $\mathbf{f}_*$  for which the optimization got trapped on the original subspace. As one would expect, tripling the dimension of the parameter space we optimize over from  $6 \times 140$  to  $6 \times 420 = 2520$  does increase the attainable fidelities slightly, but in all the cases we studied the improvements were small, and the optimization over the larger subspace exhibited trapping behavior similar to that on the original control space, terminating with substantially lower fidelities after many (time-consuming) iterations than typical successful runs (see Fig. 5). That refining the mesh would lead to higher fidelities is



**Figure 5.** Cumulative plot of the number of runs versus the final error. (1) after runs reached the  $10^{-4}$  error threshold or exceeded a thousand BFGS iterations (blue, largest), (2) after increase in fidelity dropped below  $10^{-9}$  in additional Newton iteration steps (red, intermediate), (3) after increase in fidelity obtained by a Newton iteration remains below  $5 \times 10^{-9}$  even after sub-division of time intervals (yellow, smallest). The inset focuses on the failed runs. Since the rank order of the final errors between runs did not change much across the three stages, one can see that the gains made by going beyond a thousand iterations or refining the mesh are marginal. In fact, the largest reduction in the error obtained by refining the mesh was only 7%, orders of magnitude less than what would have been achieved by a successful run with less computational effort.

rather inevitable, since even if there is a local maximum in infinite dimensions, any finite dimensional subspace will almost surely not contain this point, but the distance between increasingly larger subspaces and the point must decrease, tending to zero, allowing the fidelity to increase towards an asymptotic value. This phenomenon can conveniently be visualised by considering a non-degenerate quadratic potential on the plane and a generic line through the origin: the maximum on the line is lower than the maximum over the plane. The fact that fidelity increases so little (relative to the error) after we refine the mesh is, on the other hand, an indication that said asymptotic value is likely to be strictly below the global maximum of 1, and the data certainly suggests that the neighborhoods of the  $\mathbf{f}_*$  are really bad neighborhoods for the optimization.

## 6. Conclusion

We have revisited the control landscape for several classes of canonical quantum control problems, in particular pure-state transfer and unitary gate optimization problems.

Although the pure-state transfer fidelity as a function over the unitary group only has two types of critical points, corresponding to either the global minimum 0 or the global maximum 1, detailed analysis shows that the class of critical points for the actual optimization problem over functions in  $L^2(0, T)$  is strictly larger than the set of kinematic critical points. In particular, for any bilinear control system  $H_f(t) = H_0 + f(t)H_1$  and any fixed target time  $T$ , there exist pairs of initial and target states, such that some  $f \equiv \text{const.}$  is a critical point of the system, and we can

achieve any value of the fidelity between 0 and 1 for such critical points. Moreover, while these critical points are not expected to be attractive in most cases, we have presented examples of systems with attractive suboptimal critical points, showing that traps *do* exist for such problems, and the fidelity at these traps can take many values, unlike the fidelity for kinematic critical points, which is limited to 0 or 1 for pure-state optimization problems.

For the problem of unitary operator optimization we demonstrated that there are indeed no traps when we consider the fidelity as a functional over the unitary group. Although there are critical manifolds on which the fidelity takes values between  $-1$  (global minimum) and  $+1$  (global maximum), all of these critical points are indeed saddle points, but the situation changes when the analysis is restricted to the special unitary group  $\mathbf{SU}(N)$ . This case may appear artificial but it is actually highly relevant for quantum control, as many quantum control problems involve control Hamiltonians that have zero-trace, and hence we have no global phase control. Although the system may be  $\mathbf{U}(N)$  controllable when  $\text{Tr}[H_1] = 0$ , if the target time is fixed, so is the global phase. Perhaps surprisingly, there are more critical points in this case than for  $\mathbf{U}(N)$ , and some of these are attractive critical points at which the fidelity assumes values  $< 1$ , i.e., traps. More careful analysis shows that these traps correspond to solutions  $U = e^{i\theta}V$ , where  $e^{i\theta}$  is a root of unity, which immediately shows that these solutions have fidelity  $< 1$ , according to the standard definition of the fidelity. Nonetheless, these solutions are perfectly adequate for most practical purposes as we usually do not care about the global phase of an operator, and this problem can be avoided entirely simply by modifying the performance index to reflect the fact that we do not care about the global phase. Optimizing over the projective unitary group  $\mathbf{PU}(N)$  there are indeed no traps, at least on the group. Again, the situation is more complicated for actual optimization problem over controls in  $L^2(0, T)$ . Not only are there critical points  $f$  for which the fidelity assumes critical values other than those permitted over  $\mathbf{U}(N)$  or  $\mathbf{SU}(N)$ , but examples again show that the Hessian at these critical points can be negative definite, implying that they are locally attractive on any finite-dimensional subspace and therefore traps. The results can even be extended to the case where the target time  $T$  is variable, again proving the existence of traps.

The specific examples of traps constructed prove a theoretical point about the existence of traps, but they are not of particular interest in their own right. Unfortunately, due to the difficulty of exactly calculating the Hessian for complex systems with time-varying controls, it is difficult to *prove* the existence of traps for more interesting cases. Therefore we turned to numerical simulations to investigate any evidence for the existence of trapping regions in practice. For a three-qubit gate optimization problem we studied the simulation results suggests that there are genuine traps, i.e., attractive critical points for which the fidelity is below a certain threshold. Although such traps do not appear to be very common – in the example considered we observed trapping in about 42 out of 1000 cases, they do complicate the control landscape and obstruct convergence to a global extremum, and perhaps even more importantly, trapped runs can be extremely time-consuming, which suggests that it is important to consider the possibility of trapping in the algorithm design, to ensure that trapped runs are quickly identified and terminated. Furthermore, the example considered was a rather simple system of dimension  $N = 8$  with six independently variable controls. It is quite plausible that the number of traps and the likelihood of a local optimization algorithm converging to these increases with the system dimension or complexity. This is an important issue of practical concern that

requires a systematic further analysis.

### Acknowledgments

We would like to thank Thomas Schulte-Herbrüggen, Robert Zeier and John V. Leahy for helpful discussions and suggestions. SGS acknowledges funding from EPSRC ARF Grant EP/D07192X/1, the EPSRC QIP Interdisciplinary Research Collaboration (IRC), Hitachi and NSF Grant PHY05-51164. PdF acknowledges funding from EPSRC and Hitachi (CASE/CNA/07/47).

### Appendix A. Definition of Operators $S$ and $C$

We define the operators  $C$  and  $S$  via their action on functions  $\alpha, \beta \in L^2(0, T)$

$$\langle \beta | C \alpha \rangle_{L^2} = \int \int_{0 < \sigma < \tau < T} \cos(\omega(\sigma - \tau)) [\alpha(\tau)\beta(\sigma) + \beta(\tau)\alpha(\sigma)] d\sigma d\tau, \quad (\text{A.1a})$$

$$\langle \beta | S \alpha \rangle_{L^2} = \int \int_{0 < \sigma < \tau < T} \sin(\omega(\sigma - \tau)) [\alpha(\tau)\beta(\sigma) + \beta(\tau)\alpha(\sigma)] d\sigma d\tau \quad (\text{A.1b})$$

For  $C$ , using the trigonometric identity  $\cos(a - b) = \cos(a)\cos(b) + \sin(a)\sin(b)$ , and symmetrization the double integral, we find

$$C = \Pi[\sin(\omega \bullet)] + \Pi[\cos(\omega \bullet)] \quad (\text{A.2})$$

with  $\omega$  arbitrary. When  $\omega = \frac{\pi}{T}$ ,  $S$  can be expressed in terms of its eigenfunction decomposition, constituted of the real Fourier basis

$$S = -\frac{4}{\pi} \Pi[1] + \sum_{k=1}^{\infty} \frac{4}{\pi(4k^2 - 1)} (\Pi[\sin(2k\omega \bullet)] + \Pi[\cos(2k\omega \bullet)]) \quad (\text{A.3})$$

Recall that here  $\Pi$  denotes the unnormalized projection onto its argument and  $\bullet$  is a placeholder for the argument to an anonymous function. For example,  $\Pi[\sin(\omega \bullet)]$  is the projector onto the function  $t \mapsto \sin(\omega t)$ , i.e. the operator sending each element  $\alpha \in L^2(0, T)$  to the element  $\int_0^T \sin(\omega \tau) \alpha(\tau) d\tau \sin(\omega \bullet)$  in  $L^2(0, T)$ .

### References

- [1] Rabitz H 2009 *New Journal of Physics* **11** 105030 ISSN 1367-2630
- [2] Somloi J, Kazakov V A and Tannor D J 1993 *Chemical physics* **172** 85–98
- [3] Wang X and Schirmer S G 2009 *0906.1830* Phys. Rev. A **80**, 042305 (2009)
- [4] Nigmatullin R and Schirmer S G 2009 *New Journal of Physics* **11** 105032 ISSN 1367-2630
- [5] Chakrabarti R and Rabitz H 2007 *Int. Rev. Phys. Chem.* **26** 671–735
- [6] Rabitz H A, Hsieh M M and Rosenthal C M 2004 *Science* **303** 1998–2001
- [7] Rabitz H 2004 *J. Mod. Opt.* **51** 2469–2475 ISSN 09500340
- [8] Rabitz H, Hsieh M and Rosenthal C 2005 *Phys. Rev. A* **72** 052337
- [9] Rabitz H, Hsieh M and Rosenthal C 2006 *J. Chem. Phys.* **124** 204107–6
- [10] Rabitz H, Ho T, Hsieh M, Kosut R and Demiralp M 2006 *Phys. Rev. A* **74** 012721–9
- [11] Ho T and Rabitz H 2006 *J. Photochemistry Photobiology A: Chemistry* **180** 226–240 ISSN 1010-6030
- [12] Hsieh M and Rabitz H 2008 *Phys. Rev. A* **77** 042306–5
- [13] Hsieh M, Wu R, Rosenthal C and Rabitz H 2008 *J. Phys. B* **41** 74020

- [14] Wu R, Rabitz H and Hsieh M 2008 *J. Phys. A* **41** 15006–15100
- [15] Ho T, Dominy J and Rabitz H 2009 *Phys. Rev. A* **79** 013422–16
- [16] Montgomery R 1994 *SIAM J. Control Optim.* **32** 1605–1620
- [17] Bonnard B and Kupka I 1997 *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis* **14** 167–186 ISSN 0294-1449
- [18] Chitour Y, Jean F and Trlat E 2006 *Journal of Differential Geometry* **73** 45–73
- [19] Chitour Y, Jean F and Trelat E 2008 *SIAM Journal on Control and Optimization* **47** 1078 ISSN 03630129
- [20] Wu R, Dominy J, Ho T and Rabitz H 2009 *0907.2354*
- [21] Albertini F and D'Alessandro D 2001 *quant-ph/0106128*
- [22] Jurdjevic V 1997 *Geometric Control Theory (Chapter 3)* (Cambridge: Cambridge University Press) ISBN 0521495024
- [23] Frankel T 1965 *Differential and Combinatorial Topology: A Symposium in Honor of Marston Morse* (Princeton Univ. Press, Princeton, NJ) pp 37–53
- [24] Nocedal J and Wright S J 1999 *Numerical Optimization* (New York: Springer) ISBN 0387987932